

06 - Inference on the Mean Vector

Junvie Pailden

SIUE, F2017, Stat 589

September 07, 2017

Multivariate vs Univariate Tests

The number of parameters in multivariate tests can be large: p - means; p - variances; $\binom{p}{2}$ covariances.

Arguments for multivariate tests:

- If we use p univariate tests - it will inflate the overall Type I error rate α , whereas multivariate test preserves exact α .
- For example, if we run $p = 10$ univariate tests on the mean at $\alpha = .05$. Let $H_0^i : \mu_i = \mu_{i0}$, $i = 1, \dots, 10$.

$P(\text{overall Type 1 error})$

$$= P\{(\text{Reject } H_0^1) \cup \dots \cup (\text{Reject } H_0^{10})\}$$

$$= 1 - P\{(\text{Fail to Reject } H_0^1) \cap \dots \cap (\text{Fail to Reject } H_0^{10})\}$$

$$= 1 - 0.95^{10} = 0.40 \quad \{\text{if variables are uncorrelated - rare}\}$$

$$= \text{between } .05 \text{ and } .40 \quad \{\text{if variables are correlated}\}$$

Multivariate vs Univariate Tests

- Univariate tests completely ignore correlations.
- Multivariate tests make direct use of the correlations.
- Multivariate tests are more powerful in many cases

Power = probability reject H_0 when it is false

- Many multivariate tests involving means have as a byproduct the construction of linear combinations of variables that reveal more about how the variables unite to reject the null hypothesis.

Univariate t-test on $H_0 : \mu = \mu_0$

- We will look only at the two-tailed test because 1 tailed tests do not readily generalize to the multivariate situation.
- $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$, where σ^2 is unknown, the appropriate test statistic is

$$t = \frac{(\bar{X} - \mu_0)}{s/\sqrt{n}}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $s^2 = \frac{1}{n-1} \sum_{j=1}^n (X_i - \bar{X})^2$.

- The test statistic t has a student's t -distn with $n - 1$ degrees of freedom.
- We reject H_0 , that μ_0 is a plausible value of μ , if $|t|$ exceeds a specified percentage point of a t -distn with $n - 1$ df.

Multivariate Hotelling Test on $H_0 : \mu = \mu_0$

- The square of t (square distance) can be written as

$$t^2 = \frac{(\bar{X} - \mu_0)^2}{s^2/n} = n(\bar{X} - \mu_0)(s^2)^{-1}(\bar{X} - \mu_0)$$

- A multivariate generalization of the squared distance

$$\begin{aligned} T^2 &= (\bar{\mathbf{X}} - \mu_0)' \left(\frac{1}{n} \mathbf{S} \right)^{-1} (\bar{\mathbf{X}} - \mu_0) \\ &= n(\bar{\mathbf{X}} - \mu_0)' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \mu_0) \end{aligned}$$

- The statistic T^2 is called Hotelling's T^2 . If the observed T^2 is too large, the hypothesis $H_0 : \mu = \mu_0$ is rejected.

Hotelling Test for $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

- Under $H_0 : \mu = \mu_0$,

$$T^2 \text{ is distributed as } \frac{(n-1)p}{(n-p)} F_{p,n-p}$$

where $F_{p,n-p}$ denotes a r.v. with an F-distribution with p and $n-p$ df's.

- At the α level of significance, we reject H_0 in favor of H_1 if the observed

$$T^2 = n(\bar{\mathbf{X}} - \mu_0)' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \mu_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

where $F_{p,n-p}(\alpha)$ is the upper $(100\alpha)\text{th}$ percentile of the $F_{p,n-p}$ distribution.

Properties of Hotelling T^2 test

- $n - 1$ must be greater than p . If not, \mathbf{S} is singular and no inverse exists.
- The distribution of Hotelling's T^2 when H_0 is true and $X_i \sim N_p(\mu, \Sigma)$ has two parameters - v and p .
- In the one sample case, $v = n - 1$. In the two sample case, $v = n_1 + n_2 - 1$, where n_1 and n_2 are the sample sizes of samples 1 and 2, respectively.
- Test is always 2-sided.
- In the multivariate case, $T_{p,n-1}^2 = \frac{(n-1)p}{n-p} F_{p,n-p}$. Thus, if there are no T^2 table, we can use F tables.

Invariance of T^2 statistic

Suppose there are changes in the units of measurements for \mathbf{X} of the form

$$\mathbf{Y}_{(p \times 1)} = \mathbf{C}_{(p \times p)} \mathbf{X}_{(p \times 1)} + \mathbf{d}_{(p \times 1)}, \quad \mathbf{C} \text{ nonsingular}$$

This happens usually when variable X_i is transformed to $a_i(X_i - b_i)$, where $a_i > 0$, b_i are constants.

Given observations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, we have

$$\bar{\mathbf{y}} = \mathbf{C}\bar{\mathbf{x}} + \mathbf{d} \quad \text{and} \quad \mathbf{S}_y = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' = \mathbf{CSC}'$$

$$\mu_{\mathbf{Y}} = E(\mathbf{Y}) = E(\mathbf{CX} + \mathbf{d}) = E(\mathbf{CX}) + \mathbf{d} = \mathbf{C}\mu + \mathbf{d}$$

Invariance of T^2 statistic

Thus, \mathbf{T}^2 with the \mathbf{y}' s and a hypothesis value $\mu_{\mathbf{Y},0} = \mathbf{C}\mu_0 + \mathbf{d}$ is

$$\begin{aligned} T^2 &= n(\bar{\mathbf{y}} - \mu_{\mathbf{Y},0})' \mathbf{S}^{-1} (\bar{\mathbf{y}} - \mu_{\mathbf{Y},0}) \\ &= n(\mathbf{C}(\bar{\mathbf{x}} - \mu_0))' (\mathbf{CSC}')^{-1} \mathbf{C}(\bar{\mathbf{x}} - \mu_0) \\ &= n(\bar{\mathbf{x}} - \mu_0)' \mathbf{C}' (\mathbf{CSC}')^{-1} \mathbf{C}(\bar{\mathbf{x}} - \mu_0) \\ &= n(\bar{\mathbf{x}} - \mu_0)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu_0) \end{aligned}$$

Unrestricted Maximum MVN Likelihood and MLE's

- The maximum of the multivariate normal likelihood with no restriction on the values of μ and Σ is

$$\max_{\mu, \Sigma} L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}} \exp \left(-\frac{1}{2} \sum_{j=1}^n (\mathbf{x}_j - \hat{\mu})' \Sigma^{-1} (\mathbf{x}_j - \hat{\mu}) \right)$$

where

$$\hat{\mu} = n^{-1} \sum_{j=1}^n \mathbf{x}_j, \quad \hat{\Sigma} = n^{-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$$

are the maximum likelihood estimates (MLE) of Σ and μ , respectively.

- We can interpret $\hat{\mu}$ and $\hat{\Sigma}$ as choices for μ and Σ that best explain the observed values of the random variable.

Linear Algebra Review

(Result 4.9) Let \mathbf{A} be a $k \times k$ symmetric matrix and \mathbf{x} be a $k \times 1$ vector. Then

1. $\mathbf{x}'\mathbf{A}\mathbf{x} = tr(\mathbf{x}'\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\mathbf{x}\mathbf{x}')$
2. $tr(\mathbf{A}) = \sum_{i=1}^k \lambda_i$, where the λ_i are the eigenvalue of \mathbf{A} .

(Result 4.10) Given a $p \times p$ symmetric positive definite matrix \mathbf{B} and a scalar $b > 0$, it follows that

$$\frac{1}{|\Sigma|^b} e^{-tr(\Sigma^{-1}\mathbf{B})/2} \leq \frac{1}{|\mathbf{B}|^b} (2b)^{pb} e^{-bp}$$

for all positive definite $\Sigma_{(p \times p)}$, with equality holding only for $\Sigma = (1/2b)\mathbf{B}$.

Restricted MVN Likelihood under $H_0 : \mu = \mu_0$

- Under the hypothesis $H_0 : \mu = \mu_0$, the MVN likelihood is

$$L(\mu_0, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left(-\frac{1}{2} \sum_{j=1}^n (\mathbf{x}_j - \mu_0)' \Sigma^{-1} (\mathbf{x}_j - \mu_0) \right)$$

- The mean μ_0 is now fixed, but Σ can be varied to find the value that is “most likely” to have led, with μ_0 fixed, to the observed sample.

- This value is obtained by maximizing $L(\mu_0, \Sigma)$ with respect to Σ . Applying Result 4.9 on the exponent, we have

$$\begin{aligned} & -\frac{1}{2} \sum_{j=1}^n (\mathbf{x}_j - \mu_0)' \Sigma^{-1} (\mathbf{x}_j - \mu_0) \\ &= -\frac{1}{2} \sum_{j=1}^n \text{tr} \left[\Sigma^{-1} (\mathbf{x}_j - \mu_0) (\mathbf{x}_j - \mu_0)' \right] \\ &= -\frac{1}{2} \text{tr} \left[\Sigma^{-1} \left(\sum_{j=1}^n (\mathbf{x}_j - \mu_0) (\mathbf{x}_j - \mu_0)' \right) \right] \end{aligned}$$

Restricted MVN Maximum Likelihood under $H_0 : \mu = \mu_0$

- Let $B = \sum_{j=1}^n (\mathbf{x}_j - \mu_0)(\mathbf{x}_j - \mu_0)'$ and $b = n/2$, and applying Result 4.10, we have

$$\begin{aligned} L(\mu_0, \Sigma) &= \frac{1}{(2\pi)^{pb} |\Sigma|^b} e^{-\frac{1}{2} \text{tr}[\Sigma^{-1} \mathbf{B}]} \leq \frac{1}{(2\pi)^{pb} |\mathbf{B}|^b} (2b)^{pb} e^{-bp} \\ &= \frac{1}{(2\pi)^{np/2} |\mathbf{B}|^{n/2}} (n)^{np/2} e^{-np/2} \\ &= \frac{1}{(2\pi)^{np/2} (n^{-p} |\mathbf{B}|)^{n/2}} e^{-np/2} \\ &= \frac{1}{(2\pi)^{np/2} |\hat{\Sigma}_0|^{n/2}} e^{-np/2} \\ &= \max_{\Sigma} L(\mu_0, \Sigma) \end{aligned}$$

where $\hat{\Sigma}_0 = n^{-1} \sum_{j=1}^n (\mathbf{x}_j - \mu_0)(\mathbf{x}_j - \mu_0)'$.

Likelihood Ratio Statistic

- There is a general principle in constructing test procedures called the **likelihood ratio method**. The T^2 -statistic can be derived as the likelihood ratio test of $H_0 : \mu = \mu_0$.
- To determine whether μ_0 is a plausible value of μ , the maximum of $L(\mu_0, \Sigma)$ is compared with the unrestricted maximum of $L(\mu, \Sigma)$.

Likelihood Ratio Statistic

- The resulting ratio is called the likelihood ratio statistic

$$\begin{aligned}\text{Likelihood ratio} = \Lambda &= \frac{\max_{\Sigma} L(\mu_0, \Sigma)}{\max_{\mu, \Sigma} L(\mu, \Sigma)} \\&= \frac{\frac{1}{(2\pi)^{np/2} |\hat{\Sigma}_0|^{n/2}} e^{-np/2}}{\frac{1}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}} e^{-np/2}} \\&= \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2}\end{aligned}$$

Wilks Lambda Statistic

- The equivalent statistic $\Lambda^{2/n} = |\hat{\Sigma}|/|\hat{\Sigma}_0|$ is called the Wilk's lambda.
- The likelihood ratio test of $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ rejects H_0 if

$$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2} = \left(\frac{|\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'|}{|\sum_{j=1}^n (\mathbf{x}_j - \mu_0)(\mathbf{x}_j - \mu_0)'|} \right)^{n/2} < c_\alpha$$

where c_α is the lower $(100\alpha)th$ percentile of the distribution of Λ .

Wilks Lambda Statistic

- Determining the distribution of Λ is complicated, but, fortunately, we can write Λ in terms of the Hotelling's T^2 statistic.
- Under $H_0 : \mu = \mu_0$,

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{(n-1)} \right)^{-1}, \text{ or}$$
$$T^2 = \frac{(n-1)|\hat{\Sigma}_0|}{|\hat{\Sigma}|} - (n-1)$$

Example: Reaven and Miller (1979) link

Reaven and Miller measured five variables in a comparison of normal patients and diabetics. We use partial data for normal patients only. The three variables of major interest were

- X_1 = glucose intolerance,
- X_2 = insulin response to oral glucose,
- X_3 = insulin resistance.

The two additional variables of minor interest were

- * Y_1 = relative weight,
- * Y_2 = fasting plasma glucose.

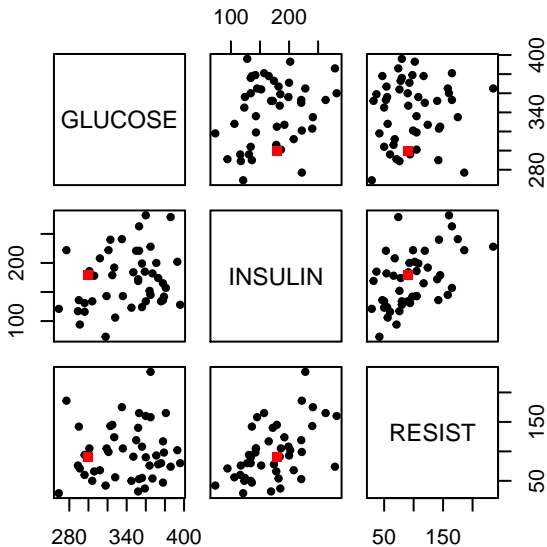
Reaven and Miller (1979): Descriptives Stats

```
patients <- read.csv("patients.csv", header = TRUE)
data.frame(Mean = colMeans(patients),
           Median = apply(patients, 2, median),
           Variance = apply(patients, 2, var))
```

#	Mean	Median	Variance
# WEIGHT	0.92	0.94	1.6e-02
# FASTING	90.41	90.00	7.1e+01
# GLUCOSE	340.83	351.00	1.1e+03
# INSULIN	171.37	170.50	2.4e+03
# RESIST	97.78	92.00	2.1e+03

```
X <- with(patients, cbind(GLUCOSE, INSULIN, RESIST))
```

$\mu_0 = (300, 180, 90)$ hypothesized mean, boundary levels



Test to see if the mean vector is different from specified vector

$$H_0 : \mu = [300, 180, 90]' \text{ vs } H_1 : \mu \neq [300, 180, 90]'$$

```
n <- nrow(X) # sample size  
p <- ncol(X) # number of variables  
(Xbar <- colMeans(X)) # sample mean
```

```
# GLUCOSE  INSULIN  RESIST  
#      341      171      98
```

```
(S <- cov(X)) # sample covariance matrix
```

```
#           GLUCOSE  INSULIN  RESIST  
# GLUCOSE    1106      397    108  
# INSULIN     397    2382    1143  
# RESIST      108    1143    2136
```

```
Sinv <- matrixcalc::matrix.inverse(S)
```

Using raw code to compute T^2

```
# hypothesized mean
mu0 <- c(300, 180, 90)
# Hotelling T2
T2 <- n*t(Xbar - mu0)%*%Sinv%*(Xbar - mu0)
# Critical Value at 5% level of significance
cval <- ((n-1)*p/(n-p))*qf(1 - 0.05, df1 = p, df2 = n-p)
# transformed T2 statistic,
# needed if compared to F crit value
T2.F <- T2/((n-1)*p/(n-p))
# Wilks Lambda
W <- (1 + T2/(n-1))^(n-2)
```

```
data.frame(T2 = T2, df1 = p, df2 = n-p,  
           Fcrit = cval, T2.trans = T2.F, Wilks = W)
```

```
#   T2 df1 df2 Fcrit T2.trans   Wilks  
# 1 87   3  43   8.9      28 1.7e-11
```

*# Since $T2 = 87 > \text{critical value} = 8.9$, then we reject H_0
at 5% level of significance.*

Using HotellingsT2() from package ICSNP

HotellingsT2() statistic is the transformed T^2 so that it has the unscaled F-distribution.

```
library(ICSNP) # install ICSNP package
(patients.T2 <- HotellingsT2(X, mu = mu0))

#
#   Hotelling's one sample T2-test
#
# data:  X
# T.2 = 30, df1 = 3, df2 = 40, p-value = 4e-10
# alternative hypothesis: true location is not equal to c(3)

patients.T2$statistic

# T.2
# 28
```